

The listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

Please amend claims 1, 6, 7, 17, 21, 22, 50, and 56 as indicated below.

Please cancel claims 9 and 24.

1. (currently amended) A method of identifying molecules for production, wherein the molecules are represented by concatenated strings, said method comprising:

i) encoding two or more related biological molecules into a data structure of initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about 10 subunits;

ii) selecting at least two substrings from said initial character strings;

iii) concatenating said substrings to form one or more product strings about the same length as one or more of the initial character strings;

iv) adding the product strings to a data structure to populate a data structure of product strings; and

v) determining sequence identities of the product strings relative to at least one initial character string; and

vi) selecting one or more product biological molecules for production, wherein the one or more product biological molecules correspond to one or more of the product strings having greater than 30% sequence identity with the at least one initial character string.

2. (previously presented) The method of claim 1, wherein said encoding comprises encoding two or more nucleic acid sequences into said character strings.

3. (previously presented) The method of claim 2, wherein said two or more nucleic acid sequences comprise a nucleic acid sequence encoding a naturally occurring protein.

4. (previously presented) The method of claim 1, wherein said encoding comprises encoding two or more amino acid sequences into said character strings.

5. (previously presented) The method of claim 4, wherein said two or more amino acid sequences comprise an amino acid sequence encoding a naturally occurring protein.

6. (currently amended) The method of claim 1, wherein said initial character strings ~~biological molecules~~ have at least 30% sequence identity with each other.

7. (currently amended) The method of claim 1, wherein said selecting in (ii) comprises selecting ~~substrings~~ at least one substring from an initial character string such that the ends of said ~~substrings~~ substring occur in string regions of about 3 to about 20 characters in the initial character string that have higher sequence identity with the corresponding region of another of said initial character strings than the overall sequence identity between the ~~same~~ two initial character strings.

8. (previously presented) The method of claim 1, wherein said selecting in (ii) comprises selecting substrings such that the ends of said substrings occur in predefined motifs of about 4 to about 8 characters.

9. (canceled)

GI
covt.
10. (previously presented) The method of claim 1, wherein said selecting in (ii) comprises aligning two or more of said initial character strings to maximize pairwise identity between two or more substrings of the initial character strings, and selecting a character that is a member of an aligned pair for the end of one of the two or more substrings.

11. (canceled)

12. (previously presented) The method of claim 1, wherein said method further comprises randomly altering one or more characters of said initial or product character strings.

13. (original) The method of claim 12, wherein said method further comprises randomly selecting and altering one or more occurrences of a particular preselected character in said character strings.

14. (previously presented) The method of claim 1, wherein said encoding, selecting, or concatenating is performed on an internet site.

15. (previously presented) The method of claim 1, wherein said encoding, selecting, or concatenating is performed on a server.

16. (previously presented) The method of claim 1, wherein said encoding, selecting, or concatenating is performed on a client linked to a network.

17. (currently amended) A computer program product on a computer readable media comprising computer code that:

- i) encodes two or more biological molecules into initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about ten subunits;
- ii) selects at least two initial substrings from said character strings;
- iii) concatenates said substrings to form one or more product strings about the same length as one or more of the initial character strings;
- iv) adds the product strings to a data structure to populate a data structure of product strings; and
- v) determines sequence identities of the product strings relative to at least one initial character string; and
- vi) selects one or more product biological molecules for production, wherein the one or more product biological molecules correspond to one or more of the product strings having greater than 30% sequence identity with the at least one initial character string.

18. (previously presented) The computer program product of claim 17, wherein said two or more biological molecules are nucleic acid sequences encoding naturally occurring proteins.

19. (previously presented) The computer program product of claim 17, wherein said two or more biological molecules are nucleic acid sequences encoding naturally occurring proteins.

20. (previously presented) The computer program product of claim 17, wherein said two or more biological molecules are amino acid sequences.

21. (currently amended) The computer program product of claim 17, wherein said initial character strings ~~biological molecules~~ have at least 30% sequence identity.

22. (currently amended) The computer program product of claim 17, wherein said computer code selects substrings in (ii) at least one substring from an initial character string such that the ends of said ~~substrings~~ substring occur in string regions of about three to about twenty characters in the initial character string that have higher sequence identity with a

corresponding region of another of said initial character strings than the overall sequence identity between the two initial character substrings.

23. (previously presented) The computer program product of claim 17, wherein said computer code selects substrings such that the ends of said substrings occur in predefined motifs of about 4 to about 8 characters.

24. (canceled)

25. (previously presented) The computer program product of claim 17, wherein the computer code selects substrings by aligning two or more of said initial character strings to maximize pairwise identity between two or more substrings of the character strings, and selecting a character that is a member of an aligned pair for the end of one substring.

26. (canceled)

27. (previously presented) The computer program product of claim 17, wherein said computer code additionally randomly alters one or more characters of said character strings.

28. (previously presented) The computer program product of claim 27, wherein said computer code additionally randomly selects and alters one or more occurrences of a particular preselected character in said character strings.

29. (previously presented) The computer program product of claim 17, wherein said computer code is stored on media selected from the group consisting of magnetic media, optical media, and optomagnetic media.

30. (previously presented) The computer program product of claim 17, wherein said computer code is in dynamic or static memory of a computer.

31-44. (canceled)

45. (previously presented) The method of claim 1, wherein the initial character strings of (i) are related.

46. (previously presented) The method of claim 1, further comprising physically screening the molecule(s) represented by the product strings for one or more desired properties.

47. (previously presented) The method of claim 1, further comprising determining a computationally predicted property for molecules represented by the product strings.

48. (previously presented) The method of claim 1, wherein the molecules represented by the product strings are made in parallel in an array of vessels.

49. (previously presented) The method of claim 1, wherein the molecules represented by the product strings are made by assembly of oligonucleotides.

50. (currently amended) The method of claim 1, further comprising testing members of the data structure of product strings for a particular property and determining sequence differences responsible for differences in the particular property ~~an optimal combination of sequences~~ using multi-variate analysis.

51. (previously presented) The computer program product of claim 17, wherein the initial character strings of (i) are related.

52. (previously presented) The computer program product of claim 17, wherein the code instructs physical screening of the molecule(s) represented by the product strings for one or more desired properties.

53. (previously presented) The computer program product of claim 17, wherein the code instructs determination of a computationally predicted property for molecules represented by the product strings.

54. (previously presented) The computer program product of claim 17, wherein the molecules represented by the product strings are made in parallel in an array of vessels.

55. (previously presented) The computer program product of claim 17, wherein the molecules represented by the product strings are made by assembly of oligonucleotides.

56. (currently amended) The computer program product of claim 17, wherein the code tests members of the data structure of product strings for a particular property and determines sequence differences responsible for differences in the particular property ~~an optimal combination of initial sequences~~ using multi-variate analysis.

57. (previously presented) A method of identifying molecules for production, wherein the molecules are represented by concatenated strings, said method comprising:

- i) encoding two or more related biological molecules into a data structure of initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about 10 subunits;
- ii) selecting at least two substrings from said initial character strings;
- iii) concatenating said substrings to form one or more product strings;
- iv) adding the product strings to a data structure to populate a data structure of product strings; and
- v) determining whether the product strings have at least a predefined measure of similarity with at least one initial character string; and
- vi) selecting one or more product biological molecules for production, wherein the one or more product biological molecules correspond to one or more of the product strings determined to have greater than the predefined value of sequence identity with at least one initial string.

G1
cont.
~~Please add the following new claims 58-131:-~~

58. (new) The method of claim 1, wherein the one or more product strings of (vi) have greater than 50% sequence identity with the at least one initial character string.

59. (new) The method of claim 1, wherein the one or more product strings of (vi) have greater than 75% sequence identity with the at least one initial character string.

60. (new) The method of claim 1, wherein the one or more product strings of (vi) have greater than 85% sequence identity with the at least one initial character string.

61. (new) The method of claim 1, wherein the one or more product strings of (vi) have greater than 90% sequence identity with the at least one initial character string.

62. (new) The method of claim 1, wherein the one or more product strings of (vi) have greater than 95% sequence identity with the at least one initial character string.

63. (new) The computer program product of claim 17, wherein the one or more product strings of (vi) having greater than 50% sequence identity with the at least one initial character string.

64. (new) The computer program product of claim 17, wherein the one or more product strings of (vi) having greater than 75% sequence identity with the at least one initial character string.

65. (new) The computer program product of claim 17, wherein the one or more product strings of (vi) having greater than 95% sequence identity with the at least one initial character string.

66. (new) A method of identifying molecules for production, wherein the molecules are represented by concatenated strings, said method comprising:

i) encoding two or more biological molecules into a data structure of initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about 10 subunits;

ii) selecting at least two substrings from said initial character strings;

iii) concatenating said substrings to form one or more product strings about the same length as one or more of the initial character strings;

iv) adding the product strings to a data structure to populate a data structure of product strings;

v) providing an alignment of the product strings; and

vi) selecting one or more product biological molecules for production, wherein the one or more product biological molecules correspond to one or more of the product strings having greater than 30% sequence identity with the at least one initial character string.

67. (new) The method of claim 66, wherein said encoding comprises encoding two or more amino acid sequences into said character strings, and wherein said two or more amino acid sequences comprise an amino acid sequence encoding a naturally occurring protein.

68. (new) The method of claim 66, wherein said initial character strings have at least 30% sequence identity with each other.

69. (new) The method of claim 66, wherein said selecting in (ii) comprises selecting at least one substring from an initial character string such that the ends of said substring occur in string regions of about 3 to about 20 characters in the initial character string that have higher sequence identity with the corresponding region of another of said initial character strings than the overall sequence identity between the two initial character strings.

70. (new) The method of claim 66, wherein said selecting in (ii) comprises selecting substrings such that the ends of said substrings occur in predefined motifs of about 4 to about 8 characters.

71. (new) The method of claim 66, wherein said selecting in (ii) comprises aligning two or more of said initial character strings to maximize pairwise identity between two or more substrings of the initial character strings, and selecting a character that is a member of an aligned pair for the end of one of the two or more substrings.

72. (new) The method of claim 66, wherein said method further comprises randomly altering one or more characters of said initial or product character strings.

73. (new) The method of claim 66, wherein the one or more product strings of (vi) having greater than 50% sequence identity with the at least one initial character string.

74. (new) The method of claim 66, wherein the one or more product strings of (vi) having greater than 75% sequence identity with the at least one initial character string.

75. (new) The method of claim 66, wherein the one or more product strings of (vi) having greater than 85% sequence identity with the at least one initial character string.

76. (new) The method of claim 66, wherein the one or more product strings of (vi) having greater than 90% sequence identity with the at least one initial character string.

77. (new) The method of claim 66, wherein the one or more product strings of (vi) having greater than 95% sequence identity with the at least one initial character string.

78. (new) A computer program product on a computer readable media comprising computer code that:

- i) encodes two or more biological molecules into initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about ten subunits;
- ii) selects at least two initial substrings from said character strings;
- iii) concatenates said substrings to form one or more product strings about the same length as one or more of the initial character strings;
- iv) adds the product strings to a data structure to populate a data structure of product strings;

- v) provides an alignment of the product strings; and
- vi) selects one or more product biological molecules for production, wherein the one or more product biological molecules correspond to one or more of the product strings having greater than 30% sequence identity with the at least one initial character string.

79. (new) The computer program product of claim 78, wherein said computer code encodes two or more amino acid sequences into said character strings, and wherein said two or more amino acid sequences comprise an amino acid sequence encoding a naturally occurring protein.

80. (new) The computer program product of claim 78, wherein said initial character strings have at least 30% sequence identity with each other.

81. (new) The computer program product of claim 78, wherein said computer code selects in (ii) at least one substring from an initial character string such that the ends of said substring occur in string regions of about three to about twenty characters in the initial character string that have higher sequence identity with a corresponding region of another of said initial character strings than the overall sequence identity between the two initial character substrings.

82. (new) The computer program product of claim 78, wherein said computer code selects in (ii) by selecting substrings such that the ends of said substrings occur in predefined motifs of about 4 to about 8 characters.

83. (new) The computer program product of claim 78, wherein said computer code selects in (ii) by aligning two or more of said initial character strings to maximize pairwise identity between two or more substrings of the initial character strings, and selecting a character that is a member of an aligned pair for the end of one of the two or more substrings.

84. (new) The computer program product of claim 78, wherein said computer code further randomly alters one or more characters of said initial or product character strings.

85. (new) The computer program product of claim 78, wherein the one or more product strings of (vi) having greater than 50% sequence identity with the at least one initial character string.

86. (new) The computer program product of claim 78, wherein the one or more product strings of (vi) having greater than 75% sequence identity with the at least one initial character string.

87. (new) The computer program product of claim 78, wherein the one or more product strings of (vi) having greater than 85% sequence identity with the at least one initial character string.

88. (new) The computer program product of claim 78, wherein the one or more product strings of (vi) having greater than 90% sequence identity with the at least one initial character string.

89. (new) The computer program product of claim 78, wherein the one or more product strings of (vi) having greater than 95% sequence identity with the at least one initial character string.

90. (new) A method of identifying molecules for production, wherein the molecules are represented by concatenated strings, said method comprising:

GI
cont.
i) encoding two or more naturally occurring biological molecules into a data structure of initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about 10 subunits;

ii) selecting at least two substrings from said initial character strings;

iii) concatenating said substrings to form one or more product strings about the same length as one or more of the initial character strings;

iv) adding the product strings to a data structure to populate a data structure of product strings; and

v) selecting one or more product biological molecules for production, wherein the one or more product biological molecules correspond to one or more of the product strings having greater than 30% sequence identity with the at least one initial character string.

91. (new) The method of claim 90, wherein said encoding comprises encoding two or more nucleic acid sequences into said character strings.

92. (new) The method of claim 90, wherein said encoding comprises encoding two or more amino acid sequences into said character strings, and wherein said two or more amino acid sequences comprise an amino acid sequence encoding a naturally occurring protein.

93. (new) The method of claim 90, wherein said initial character strings have at least 30% sequence identity with each other.

94. (new) The method of claim 90, wherein said selecting in (ii) comprises selecting at least one substring from an initial character string such that the ends of said substring occur in string regions of about 3 to about 20 characters in the initial character string that have higher sequence identity with the corresponding region of another of said initial character strings than the overall sequence identity between the two initial character strings.

95. (new) The method of claim 90, wherein said selecting in (ii) comprises selecting substrings such that the ends of said substrings occur in predefined motifs of about 4 to about 8 characters.

96. (new) The method of claim 90, wherein said selecting in (ii) comprises aligning two or more of said initial character strings to maximize pairwise identity between two or more substrings of the initial character strings, and selecting a character that is a member of an aligned pair for the end of one of the two or more substrings.

97. (new) The method of claim 90, wherein said method further comprises randomly altering one or more characters of said initial or product character strings.

98. (new) The method of claim 90, wherein the one or more product strings of (v) having greater than 50% sequence identity with the at least one initial character string.

99. (new) The method of claim 90, wherein the one or more product strings of (v) having greater than 75% sequence identity with the at least one initial character string.

100. (new) The method of claim 90, wherein the one or more product strings of (v) having greater than 85% sequence identity with the at least one initial character string.

101. (new) The method of claim 90, wherein the one or more product strings of (v) having greater than 90% sequence identity with the at least one initial character string.

102. (new) The method of claim 90, wherein the one or more product strings of (v) having greater than 95% sequence identity with the at least one initial character string.

103. (new) A computer program product on a computer readable media comprising computer code that:

i) encodes two or more naturally occurring biological molecules into initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about ten subunits;

- ii) selects at least two initial substrings from said character strings;
- iii) concatenates said substrings to form one or more product strings about the same length as one or more of the initial character strings;
- iv) adds the product strings to a data structure to populate a data structure of product strings; and
- v) selects one or more product biological molecules for production, wherein the one or more product biological molecules correspond to one or more of the product strings having greater than 30% sequence identity with the at least one initial character string.

104. (new) The computer program product of claim 103, wherein said computer code encodes by encoding two or more nucleic acid sequences into said character strings.

105. (new) The computer program product of claim 103, wherein said computer code encodes two or more amino acid sequences into said character strings, and wherein said two or more amino acid sequences comprise an amino acid sequence encoding a naturally occurring protein.

106. (new) The computer program product of claim 103, wherein said initial character strings have at least 30% sequence identity with each other.

107. (new) The computer program product of claim 103, wherein said computer code selects in (ii) at least one substring from an initial character string such that the ends of said substring occur in string regions of about three to about twenty characters in the initial character string that have higher sequence identity with a corresponding region of another of said initial character strings than the overall sequence identity between the two initial character substrings.

108. (new) The computer program product of claim 103, wherein said computer code selects in (ii) by selecting substrings such that the ends of said substrings occur in predefined motifs of about 4 to about 8 characters.

109. (new) The computer program product of claim 103, wherein said computer code selects in (ii) by aligning two or more of said initial character strings to maximize pairwise identity between two or more substrings of the initial character strings, and selecting a character that is a member of an aligned pair for the end of one of the two or more substrings.

110. (new) The computer program product of claim 103, wherein said computer code further randomly alters one or more characters of said initial or product character strings.

111. (new) The computer program product of claim 103, wherein the one or more product strings of (v) having greater than 50% sequence identity with the at least one initial character string.

112. (new) The computer program product of claim 103, wherein the one or more product strings of (v) having greater than 75% sequence identity with the at least one initial character string.

113. (new) The computer program product of claim 103, wherein the one or more product strings of (v) having greater than 85% sequence identity with the at least one initial character string.

114. (new) The computer program product of claim 103, wherein the one or more product strings of (v) having greater than 90% sequence identity with the at least one initial character string.

61
C.D. 4.
115. (new) The computer program product of claim 103, wherein the one or more product strings of (v) having greater than 95% sequence identity with the at least one initial character string.

116. (new) A method of identifying molecules for production, wherein the molecules are represented by concatenated strings, said method comprising:

i) encoding two or more biological molecules into a data structure of initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about 10 subunits;

ii) selecting at least two substrings from said initial character strings;

iii) concatenating said substrings to form one or more product strings about the same length as one or more of the initial character strings;

iv) adding the product strings to a data structure to populate a data structure of product strings;

v) obtaining one or more computationally predicted properties for the product strings in the data structure; and

vi) selecting one or more product biological molecules for production on the basis of the one or more computationally predicted properties.

117. (new) The method of claim 116, wherein the computationally predicted properties comprise one or more of a maximum or minimum molecular weight, a maximum or minimum free energy, a maximum or minimum contact surface with a target molecule or surface, a specified net charge, a predicted pK, a predicted pI, a binding avidity, secondary form, and tertiary form.

118. (new) The method of claim 116, wherein said encoding comprises encoding two or more amino acid sequences into said character strings.

119. (new) The method of claim 116, wherein said selecting in (ii) comprises aligning two or more of said initial character strings to maximize pairwise identity between two or more substrings of the initial character strings, and selecting a character that is a member of an aligned pair for the end of one of the two or more substrings.

120. (new) The method of claim 116, wherein said method further comprises randomly altering one or more characters of said initial or product character strings.

GI
cont.
121. (new) The method of claim 116, wherein the one or more product biological molecules of (vi) having greater than 50% sequence identity with the at least one initial character string.

122. (new) The method of claim 116, wherein the one or more product biological molecules of (vi) having greater than 75% sequence identity with the at least one initial character string.

123. (new) The method of claim 116, wherein the one or more product biological molecules of (vi) having greater than 90% sequence identity with the at least one initial character string.

124. (new) A computer program product on a computer readable media comprising computer code that:

i) encodes two or more biological molecules into initial character strings to provide a collection of two or more different initial character strings wherein each of said biological molecules comprises at least about ten subunits;

ii) selects at least two initial substrings from said character strings;

iii) concatenates said substrings to form one or more product strings about the same length as one or more of the initial character strings;

iv) adds the product strings to a data structure to populate a data structure of product strings;

v) obtains one or more computationally predicted properties for the product strings in the data structure; and

vi) selects one or more product biological molecules for production on the basis of the one or more computationally predicted properties.

125. (new) The computer program product of claim 124, wherein the computationally predicted properties comprise one or more of a maximum or minimum molecular weight, a maximum or minimum free energy, a maximum or minimum contact surface with a target molecule or surface, a specified net charge, a predicted pK, a predicted pI, a binding avidity, secondary form, and tertiary form.

126. (new) The computer program product of claim 124, wherein the computer code encodes in (i) by encoding two or more amino acid sequences into said character strings.

127. (new) The computer program product of claim 124, wherein the computer code selects in (ii) by aligning two or more of said initial character strings to maximize pairwise identity between two or more substrings of the initial character strings, and selecting a character that is a member of an aligned pair for the end of one of the two or more substrings.

128. (new) The computer program product of claim 124, wherein the computer code further randomly alters one or more characters of said initial or product character strings.

129. (new) The computer program product of claim 124, wherein the one or more product biological molecules of (vi) having greater than 50% sequence identity with the at least one initial character string.

130. (new) The computer program product of claim 124, wherein the one or more product biological molecules of (vi) having greater than 75% sequence identity with the at least one initial character string.

131. (new) The computer program product of claim 124, wherein the one or more product biological molecules of (vi) having greater than 90% sequence identity with the at least one initial character string.